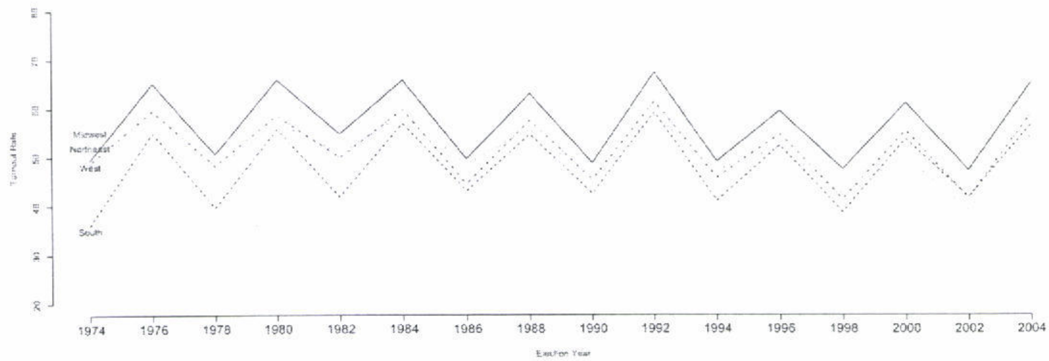

Figure 2: Estimated Turnout of Registered Voters by Region, 1974-2004. Estimates from United States Election Project (`http://elections.gmu.edu/`).

For example, if we were to look at only 2004, we might conclude that stricter voter identification requirements cause voters to turnout at lower rates because of the correlation between regional turnout rates and likelihood of adopting a more stringent identification requirement. Similarly, if we were to look at one state over time, we might make the same false inferences because of the cyclical turnout rates apparent in the graph. Consider, for example, if we were to compare a state that adopted more stringent requirement in 2002. If we compare 2000 to 2002, we would incorrectly conclude that the decline was caused by the change in identification requirements, but all states saw a drop in turnout because 2002 was a midterm election. Again, this is a critical flaw in earlier studies — by focusing solely on single presidential elections, they are confusing voter identification requirements with other causal factors that cannot be separated in the use of only a single election in their analysis.[10]

Our estimation strategy exploits the temporal and geographic variability in voter identification requirements to sidestep the problem on non-random assignment. This is referred to as a difference-in-differences estimator and our analysis is built on a generalization of this procedure. In particular, we use a multilevel model — also referred to as a random effects model — to assess how voter identification requirements affect participation by registered voters, using data from four years of recent CPS Voter Supplement data. While multilevel models have seen many applications in fields outside of political science, only in relatively recent years have we seen the use of multilevel models in political science applications and journals (e.g., Steenbergen and Jones 2002; Raudenbush and Bryk 2002; Western 1998).[11] The multilevel model allows us to control for the constant factors that cause turnout rates to vary within states and for the cyclical changes in

[10]In general, it is only possible to identify a causal effect in a single cross-section (i.e., one year's data) with random assignment or with an instrumental variable approach (Moffitt 1991).

[11]More recently, a special issue of *Political Analysis* was devoted to the topic of multilevel modeling in political methodology, with applications to a wide variety of important substantive problems (Kedar and Shively 2005).

turnout over time.

In addition to using a much richer dataset than previous studies with a generalization of a difference-in-differences estimator to minimize the problem of non-random assignment, we also attempt to handle the sparse and ordinal nature of the data. The data is sparse because with eight different types of identification requirements and only fifty states, we do not observe that many elections under a given type of procedure. The standard approach around this problem is to assume some sort of linear (or other parametric) effect. That is, if we consider our list presented at the beginning of the section, we would assume that the effect of a signature match was three times that of merely stating one's name on an individual's probability of voting, since it is third on the list. While the ordering of the list seems plausible, the linear growth (or dose-response curve) is a very strong assumption that seems implausible. We, instead, leverage the ordinal nature of the data to allow for deviations for this linear effect insofar as the data suggest via a Bayesian shrinkage estimator.

In the next section, we present the results from the aggregate component of our multilevel model, examining how voter identification requirements may affect voter participation at the state level. That is followed by a presentation of the results from our individual-level model of participation.

4. ESTIMATES FROM AGGREGATE LEVEL DATA

In addition to the unobservable effects on voter turnout, such as regional trends or yearly shocks, we want to control for any observable characteristics that might affect turnout as well. There are two approaches we consider — aggregate and individual-level data — and our model allows us to consider both levels of data simultaneously. Aggregate data can be a useful source of information about voter turnout mainly because there is no concern that survey respondents are "incorrectly remembering" turning out to vote. We know from surveys that have validated turnout of survey respondents using public voting records, misreporting occurs between five and ten percent of validated cases.[12] The use of aggregated data to study individual behavior, however, also raises concerns about aggregation bias. That is, it is not possible to draw conclusions about individual voter's decisions based solely on the analysis of aggregate data. Further, we are also interested in the impact of these identification requirements on sub-populations, such as racial and ethnic minorities and seniors. Given the coarse nature of state-level data, we can not say anything about these populations of interest.

For the aggregate analysis, following the previous literature on turnout, we gathered data on demographic variables at the state-level, such as the percentage of the population who have graduated from high school, the percentage of the population who are

[12] There are an array of published studies that have looked at the validated turnout data. See, for example, the early studies by Abramson and Claggett (1984, 1986, 1989, 1991 and 1992), or the more recent analyses by Bernstein, Chadha and Montjoy (2001) or Cassel (2004).

minorities, the unemployment rate and per capita income. The specific empirical model of voter identification requirements on state-level turnout rates for this data is:

$$\ln(\text{turnout rate}) = \alpha ID_{st} + \beta^0 + \beta^1 X_{st} + \epsilon_{st}; \quad s = 1, \ldots, 51; \quad t = 1, \ldots, 4;$$

where s indexes states and t indexes years. That is, the logarithm of the turnout rate is a linear function of observable regressors.

The turnout rate is measured relative to registered voters in the state, and the variable of interest, ID_{st}, is coded as an ordinal variable ranging from zero (state name) to seven (photo identification).[13] The vector of covariates, X_{it}, includes the following:

% HS Grad: the percent of high school graduates in state s at year t, according to the Census Bureau;
Per capita income: the per capita income in state s at year t according to the Bureau of Economic Analysis;
Unemp rate: the unemployment rate in state s at year y according to the Bureau of Economic Analysis;
South: an indicator equal to one if the state is southern and zero otherwise;
% Nonwhite: the percent of individuals in state s at year t that are reportedly not white, according to the Census Bureau.

As the level of turnout in a state may vary due to yearly shocks or regional trends (see Figure 2), random effects are included for state and year.

$$\begin{aligned} \beta^0 &= \gamma_s^0 + \gamma_t^1; \\ \gamma_s^0 &\overset{\text{iid}}{\sim} N(0, \sigma_{\gamma_s}); \\ \gamma_t^0 &\overset{\text{iid}}{\sim} N(0, \sigma_{\gamma_t}); \\ \text{for } s &= 1, \ldots, S \text{ and } t = 1, \ldots, T. \end{aligned}$$

Our results from the aggregate model can be found in Figure 3. The figure displays the estimated percentage change in turnout among registered voters at the state-level, for voter identification requirements and other contributing factors to aggregate turnout. The circles represent the point estimates, the heavy black lines denote the 50% confidence interval, and the thin black lines denote the 95% confidence interval.

As can be seen from the figure, the only state characteristic that has an effect on the rate of turnout in a state is the percentage of high school graduates. As the percentage of high school graduates in a state increases by one unit, the rate of turnout in that state increases about two percent. The coefficient on the rate of unemployment has the expected sign — increasing the unemployment rate one point has a negative impact

[13]Given the limited amount of state-level observations, it is not possible to do much more than a linear treatment specification. When we turn to the individual level specification with much more data we can relax this assumption.

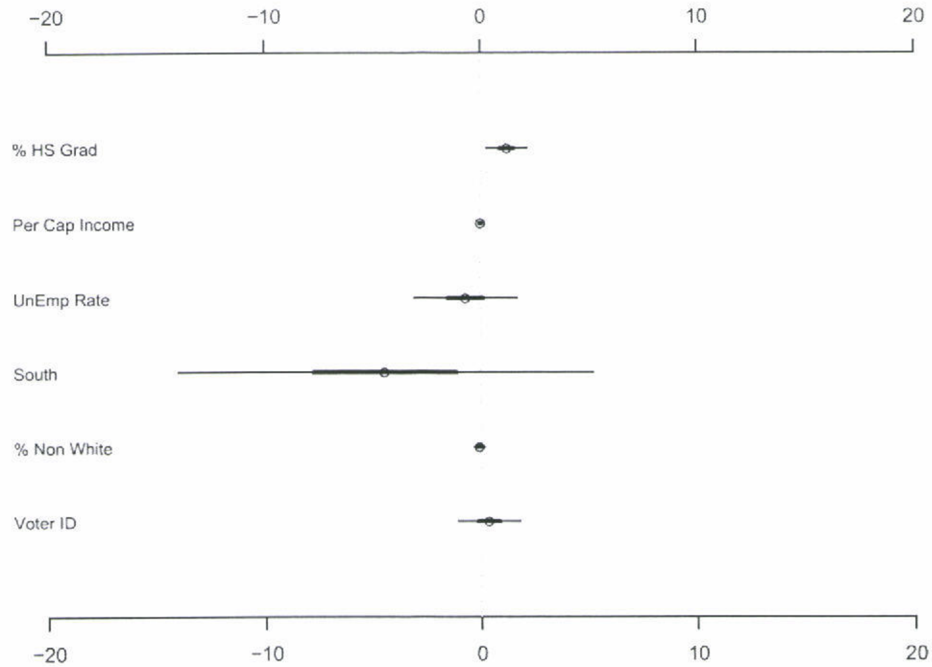

Figure 3: *Estimates of the determinants of state level turnout of registered voters, 2000-2006. The graph shows the result of regression of log(turnout) on the covariates, including state and year effects. The center dots correspond to the point estimates, the thicker lines to the 50% confidence interval, and the thinner lines to 95% confidence interval.*

on turnout — but it is not statistically significant. The other variables in the model per capita income, percent of the population nonwhite, whether the state is in the south, and interestingly voter identification — have no discernible impact on statewide turnout rates. Thus, our analysis of our critical hypothesis at the aggregate level yields no support for the claim that voter identification requirements have any effect on the turnout decisions of registered voters.

5. ESTIMATES FROM INDIVIDUAL LEVEL DATA

The aggregate data poses two problems. First, with only 50 states and four years of data, there is very little information available to inform us about the effects of voter identification requirements. Second, answering questions about voter identification laws effects on subgroups is not possible. Because we are most interested in the effect of voter identification laws on individual subgroups of voters — not on state-wide trends — and we would like to be able to more precisely identify these effects, we turn to individual

responses from the Current Population Survey in 2000, 2002, 2004 and 2006.

The CPS has a very large sample size (120,000 per year), which affords us good coverage of both states and populations of interest. We do need to worry about overreporting of turnout, an issue we return to in our discussion of future research. In addition to answering questions about voter registration and turnout, respondents to the CPS provide basic demographic information, such as their education level, age, income, sex and race. Not only do we use these demographic questions to control for varying propensities of turnout based on individual characteristics, we also are able to determine whether voter identification requirements are affecting certain groups disproportionately after controlling for other factors.

As mentioned previously, an additional complication arises because of the non-binary nature of the voter identification requirement. We could not do much about this in the aggregate level model, but with the greater number of respondents in the individual-level data we have some flexibility estimating the effects of the varying identification requirements. But given the sparseness of the data, precisely estimating individual effects for each of the eight identification requirements is difficult. This would involve coding each voter identification regime as a binary indicator variable in our model, but the concern then is that we simply will have too little information for some of the less-used regimes to identify (let alone precisely estimate) the effects of each voter identification requirement relative to the others. On the other hand, we could assume that the effect is linear across the eight requirements, as we did in the aggregate level model. That is, the effect on the probability that a voter turns out is the *same* if we change the requirement from stating one's name to signing one's name as if we change the requirement from merely requesting a photo identification card to requiring a photo identification card.[14] This is a strict assumption. As compromise we, in effect, compute a weighted average of these two approaches with the weights being proportional to the amount of information in the data about that particular identification regime.[15]

Specifically, we start with a logistic model of turnout from the CPS. Because we are interested in the effect of identification requirements *at the polls* and not the various unobserved barriers to voting associated with the registration process, the estimation is conditioned on the subset of respondents who are registered to vote. Our logistic model takes the form:

$$\Pr(Y_{it} = 1) = \text{logit}^{-1}(\alpha_{j[i]} + \beta^0 + \beta^1 X_{it}),$$
$$\text{for } j = 1, \ldots, 8; \quad i = 1, \ldots, N; \text{ and } t = 1, \ldots, 4.$$

where j indexes identification regime, i indexes the respondents, and t indexes years. The variable Y_{it} is binary and equal to one if the respondent reported voting in that year's

[14] Recall that requesting a photo identification card allows the voter the option of signing an affidavit swearing their identity and then casting a *regular* ballot, whereas requiring a photo ID only allows the voter the option of casting a *provisional* ballot.

[15] The particular analysis we use, a Bayesian shrinkage estimator, is documented in Alvarez, Bailey and Katz (2007).

election. The variable β^0 is an intercept term. The vector of covariates, X_{it}, includes the following:

South: an indicator equal to unity if the respondent resides in a southern state;
Female: an indicator equal to unity if the respondent is female;
Education: a ordinal variable indicating the reported level of education — 'some high school,' 'high school graduate,' 'some college,' or 'college graduate';
Education^2: the squared value of *Education*;
Age: the respondent's age in years;
Age^2: the squared value of *Age*;
Income: an ordinal variable indicating the reported level of household family income that takes on 13 values — ranging from 'Less than \$5,000' to 'More than \$75,000';
Non-White: an indicator equal to unity if the respondent reported a race other than White.

This covariate vector replicates what we consider to the be canonical model of voter turnout in the literature that uses CPS Voter Supplement data (e.g., Nagler 1991).

As the level of turnout in a state may vary due to yearly shocks or regional trends, random effects are included for state and year.

$$\beta^0 = \gamma^0_{s[i]} + \gamma^1_{t[i]};$$
$$\gamma^0_{s[i]} \overset{\text{iid}}{\sim} N(0, \sigma_{\gamma_s});$$
$$\gamma^0_{t[i]} \overset{\text{iid}}{\sim} N(0, \sigma_{\gamma_t});$$
$$\text{for } s = 1, \ldots, S \text{ and } t = 1, \ldots, T.$$

That is, each individual i in state s and year t share a common intercept term, with each level of intercepts pooled toward zero and with common variance.

As noted above, we could model the impact of the variable of interest, $VoterID$, as an unpooled additive effect (e.g., indicator variables for each regime), or alternatively, constrain the effect to be linear. Rather than commit to either extreme, we effectively combine the first two approaches into a sort of weighted average, where the weighting variable is determined by the data:

$$\alpha_{j[i]} = \alpha^0 + \alpha^1 ID_{it} + \nu_j,$$
$$\nu_j \overset{\text{iid}}{\sim} N(0, \sigma_\alpha).$$

That is, for each identification requirement level, j, the estimated impact on turnout is a random intercept term, ν_j, and is pooled toward a group linear impact, $\alpha^0 + \alpha^1 ID_{it}$.[16]

[16] A final consideration in the third model is interpretation of the α^0 and α^1 parameters. These parameters are partially unidentified between the linear trend in the ν_j parameters. The identification is partial, as the ν_j parameters are pooled toward zero, but with only $J = 8$ groups, converging the algorithm is time consuming. To correct for this problem, after estimation, the data is "post-processed" to obtain finite population slope parameters based on the regression of α_j on ID_j. This is equivalent to constraining the ν_j parameters to have mean zero and slope zero (Gelman and Hill, 2006).

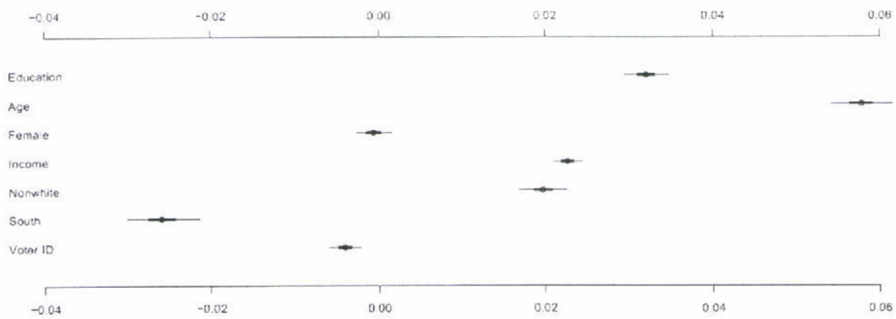

Figure 4: *Estimates of the determinants of individual level turnout of registered voters, 2000-2006. The graph shows the result of logistic regression of the probability of voting on the covariates, including state and year effects. The center dots correspond to the point estimates, the thicker lines to the 50% confidence interval, and the thinner lines to 95% confidence interval.*

Our results for the individual model can be found in Figure 4. The figure displays the estimated percentage change in the probability of turning out to vote, conditional upon being registered, for voter identification requirements and demographic control variables. The circles represent the point estimates, the heavy black lines denote the 50% confidence interval, and the thin black lines denote the 95% confidence interval.

Living in the South decreases the odds that an individual in our sample votes, while being older, more educated and wealthier increases the odds an individual turns out to vote. In our sample, being female does not effect the probability of voting, but being a minority increases the probability of turning out to vote, conditional on being registered to vote. These effects are all consistent with the previous literature on turnout, lending credence to our model's specification (e.g., Nagler 1991). Increasing the strength of voter identification requirements, on average, decreases the probability of turning out to vote. We examine the deviations from this linear trend below.

Figure 5 plots the marginal effect of voter identification regimes on the probability that a respondent turns out to vote. The horizontal axis represents the voter identification requirements. The vertical axis plots the probability of turning out to vote. The line represents the probability of voting for a mean respondent in our sample, for each identification requirement being in place. This average individual is a white male, aged 48, with some college education, has an income of between $35,000 and $39,999, and lives in Ohio in 2004. The points on the graph denote the deviation from the linear trend estimated for each requirement and the vertical bars denote the 95% intervals of uncertainty around each. Interestingly, we see that the requirements for signature matching, requiring an identification card and requiring a photo identification card have a more negative effect on participation than suggested by the simple linear model. Requesting identification cards and requesting photo identification cards is less strict than suggested by the linear

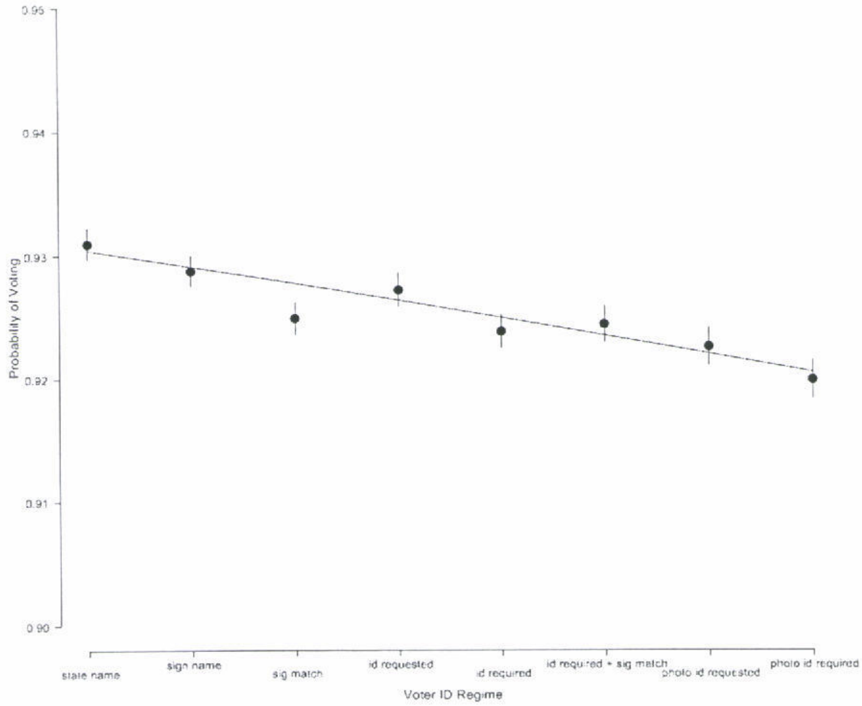

Figure 5: *Estimated probability of voting by identification requirement. The graph plots the impact for an average registered voter from the Current Population Survey (2000-2006). The estimates come from a logistic regression of the probability of voting controlling for demographic characteristics. The solid line is the linear trend that the identification effects are shrunk towards. The dots are the point estimates and the bars represent the 95% confidence intervals for the effect.*

trend. These estimates first indicate that indeed, voter identification requirements do not have a simple linear effect on the likelihood that a voter participates. In addition, we see that the stricter requirements requirements more than merely presenting a non-photo identification card — are significant negative burdens on voters, relative to a weaker requirement, such as merely signing a poll-book.

Previous studies that we are aware of, however, did not use multiple election cycles in their analysis; thus those studies have likely confused the possible effects of new voter identification requirements with the cross-sectional correlations we discussed earlier. Again, there we saw that states with low turnout were also states which had imposed strict voter identification requirements in 2004. Here, as we have data that varies by state and time, we are able to separately identify and estimate the effects of voter identification requirements on voter turnout, that is, separately from the confounding effects of past voter participation rates and voter identification regimes.

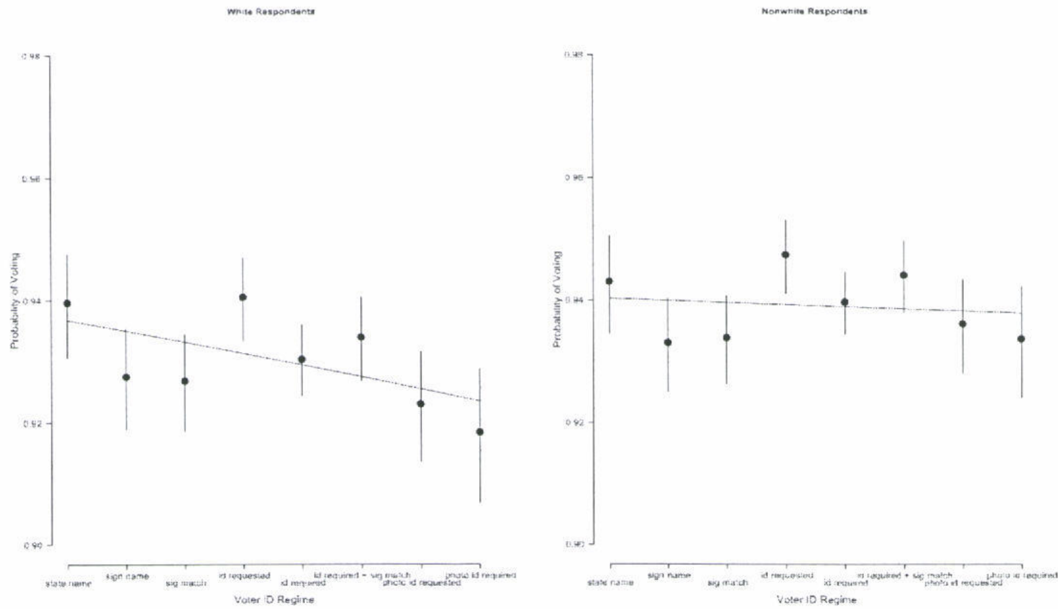

Figure 6: *Estimated probability of voting by identification requirement and race. The graphs plot the impact for an average registered voter from the Current Population Survey (2000-2006) for Whites and Non-Whites. The estimates come from a logistic regression of the probability of voting controlling for demographic characteristics. The solid line is the linear trend that the identification effects are shrunk towards. The dots are the point estimates and the bars represent the 95% confidence intervals for the effect.*

Next, we turn to the critical question of the possible interaction between the various voter identification regimes and the racial identify of registered voters in the CPS samples: do voter identification requirements, especially the stricter ones, depress the likelihood of turnout more for nonwhite registered voters than for white registered voters? To answer this question, we estimated a slight variant of the model used above, which includes interaction terms for voter identification requirements and the racial identity of the voters in the CPS samples. This model includes both the shrinkage estimator and in our linear term an interaction between the voter identification regimes and the racial identity of each registered voter. These results from this analysis are given in Figure 6.

In the left panel we give the results of the voter identification regimes for white registered voters, and in the right panel of Figure 6 the results for nonwhite registered voters. It is clear from comparison of the two graphs that we can reject the hypothesis that there is a substantial racial difference in the impact of voter identification requirements. First off, we see that the slopes differ in the two panels, and in fact, the slope for white registered voters is more strongly sloped than for nonwhite registered voters. Also, when we look at some of the specific regimes, especially the most restrictive ones, our analysis here indicates that they have a more strongly negative effect on the participation of white, relative to nonwhite voters, all other variables held constant in our model. This is an

important result. Controlling for the factors usually seen in models of voter participation, we see no evidence that voter identification requirements are racially discriminatory.

Next we turn to three other important socio-demographic variables in our turnout model: education, age and income. We are interested here in seeing whether these variables have any interactive effect with identification requirements. In particular, there is reason to believe that registered voters who are of lower educational attainment, lower income, or who are younger may more likely to be deterred from voting as identification requirements increase. These next figures plot the probability of voting conditional on being a mean respondent under each identification regime, tracing out the likelihood for voting as education, age, and income levels vary. The estimated models include an interaction term between the demographic variable of interest and identification type. The dashed lines are the confidence intervals for the random effects term only, and do not include the uncertainty in the estimate; these are provided for convenience only.

Beginning with the interactive effect between voter identification regime and educational attainment in Figure 7, we see that there is a slight, but significant, interaction between these two variables, controlling for everything else in our model. As we move from the less to more restrictive voter identification requirements, we do see that registered voters at the lower end of the educational attainment scale are less likely to participate. For example, in states that require only that a registered voter provide their name, or sign their names, relative to states that require that a registered voter produce a photographic identification, registered voters with only some high school are significantly less like to vote.

Next, in Figure 8, we see little interaction between voter identification requirements and age. In particular, we expected to see that younger voters would be significantly less likely to vote in states with more restrictive identification regimes in place; we see little evidence in this figure to support that hypothesis. Nor do we see older voters being deterred more.

Finally, we show the interaction between the other measure of socioeconomic status and voter identification regimes in Figure 9. The various panels show the different voter identification regime effects for the various levels of household family income. As we have seen in the graph in Figure 7 for education, we do see evidence of an interactive effect, even after controlling for all of the other variables in our model. This is not surprising given that household income and education levels are highly correlated. As expected, voters with lower levels of income are less likely to vote under the more restrictive voter identification regimes; comparing again the extremes of states that simply require the voter provide their name, to those states that require a photographic identification from the registered voter in order to cast a ballot, we see that lower income registered voters in the latter type of state are significantly less likely to vote.

In conclusion, our analysis of the individual-level component of our multi-year and multilevel model, we have found a number of significant results. First, we see that there is evidence to support the claim that the most restrictive forms of voter identification

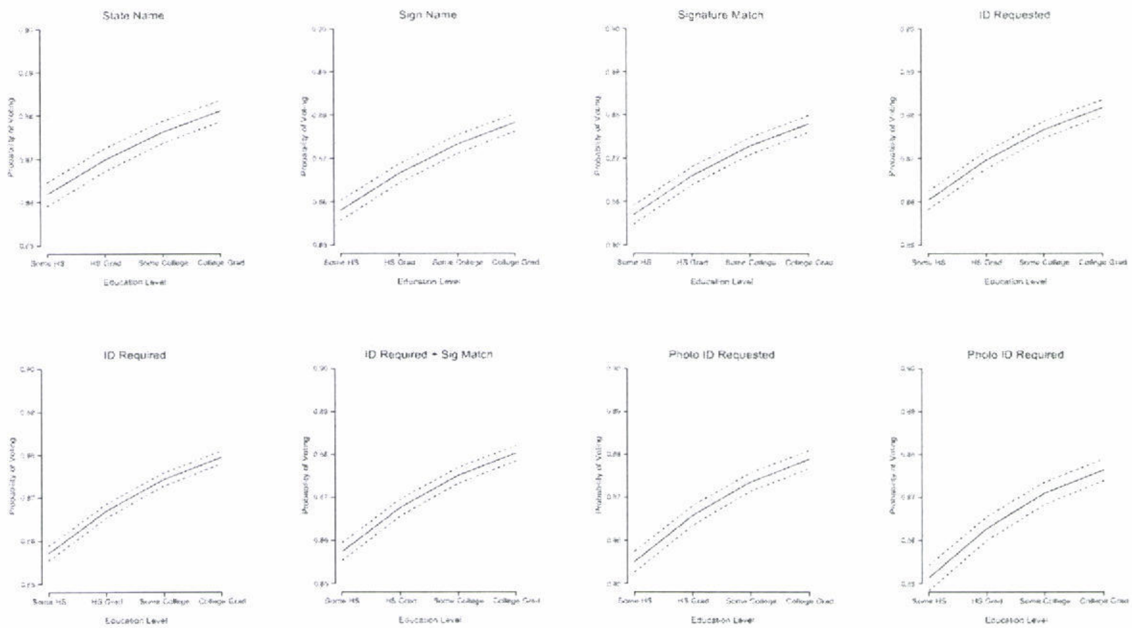

Figure 7: *Estimated probability of voting by identification requirement and education level. The graphs plot the estimated probability of voting by an average registered voter from the Current Population Survey (2000-2006) given different voter identification regimes as education levels vary. The estimates come from a logistic regression of the probability of voting controlling for demographic characteristics. The dashed lines are the confidence intervals for the random effects term only, and do not include the uncertainty in the estimate; these are provided for convenience only.*

requirements do lead to lower levels of participation by registered voters. However, we find no evidence to support the hypothesis that this effect is more profound for nonwhite registered voters, controlling for other variables, especially income and education. Yet we find that these other socioeconomic status variables, especially education and income, do show a significant interactive effect with stricter identification requirements. In particular, we find that registered voters with low levels of educational attainment or lower levels of income are less likely to vote the more restrictive the voter identification regime.

6. DISCUSSION

In general, there is scant research on the effect that voter identification requirements, of any form, have on the participation of registered voters. In an attempt to understand whether the requirements imposed by both HAVA and nearly half the states reduce registered voter participation, we used a novel methodology to study the effects of voter identification requirements on the likelihood that voters participate in these two presi-